**Modulo di Proposta Argomento Tesi**

*(compilato da un docente del Consiglio Didattico per proporre un nuovo argomento di tesi e inviato alla Commissione Tesi all’indirizzo email* [*commissionetesi.diem@unisa.it*](mailto:commissionetesi.diem@unisa.it) *)*

Eventuali tesi che non rispettano tale formato saranno scartate senza essere prese in considerazione.

**Docente relatore**: Giuseppe D’Aniello, Sabrina Senatore

**Insegnamento di riferimento**: Situation Awareness

**Titolo argomento di tesi**: **Costruzione e specializzazione di un Knowledge Graph per la Cyber Situation Awareness**

**Descrizione dell’attività**

***Premessa, breve descrizione del contesto*** *(max 10 righe)*

La Cyber Situation Awareness (Cyber SA) è cruciale per comprendere, monitorare e proiettare potenziali minacce nel contesto della sicurezza informatica.Tuttavia, le informazioni di valore spesso risiedono in report non strutturati, per cui è fondamentale l'estrazione di tali dati in tempo reale. D’altro canto, organizzare in modo strutturato tali dati può essere fondamentale per migliorare la comprensione situazionale e supportare la proiezione di possibili sviluppi futuri.

La costruzione di un Knowledge Graph (KG) per strutturare e rappresentare le informazioni provenienti da report relativi a cyber thread può migliorare significativamente la capacità di monitorare le minacce in modo da identificare le entità coinvolte e recuperare le informazioni pertenenti dalla conoscenza nel KG e rispondere prontamente a minacce e garantire la sicurezza.

***Descrizione delle metodologie e delle tecnologie che lo studente utilizzerà nell’ambito di questa tesi*** *(10-30 righe)*

Lo studente dovrà impiegare metodologie di elaborazione del linguaggio naturale (NLP), tecnologie di knowledge representation e estrarre le risposte appropriate dal Knowledge Graph.

- **Open Information Extraction**: Modelli di estrazione delle informazioni aperti per individuare relazioni ed entità (sottoforma di triple RDF) nei report.

- **Named Entity Recognition (NER)**: Modelli di riconoscimento delle entità, specializzati nel dominio della cybersecurity (ad esempio, indicatori di compromissione), per identificare termini della cybersecurity.

- **Knowledge Graph Construction**: partendo da un modello semantico-ontologico, la generazione della Knowledge Graph che integri entità e relazioni estratte da report relativi a cyber thread.

- **Linguaggi di interrogazione avanzati**: Utilizzo di linguaggi come SPARQL o Cypher per formulare query complesse e interrogare il KG in modo efficiente.

- **Supporto alle domande in linguaggio naturale**: Implementazione di meccanismi che permettano di tradurre domande in linguaggio naturale in query strutturate, ad esempio mediante il solo uso di large Language model per la traduzione in query SPARQL o Cypher.

***Descrizione dei dataset o del setup di test e di eventuali attrezzature (macchine con GPU, server ecc) che lo studente avrà a disposizione per svolgere la sua tesi*** *(10-20 righe)*

Per l’estrazione delle entità e delle relazioni, lo studente utilizzerà dataset di report relativi ad attacchi avanzati e persistenti (APT) non strutturati, come quelli pubblicati da piattaforme di threat intelligence o società di sicurezza informatica (ad es. Mandiant, CrowdStrike). Questi report contengono descrizioni dettagliate delle campagne di attacco, Indicatori di Compromissione (IoC), e profili degli attaccanti.

***Descrizione dettagliata delle attività che in tale contesto deve svolgere il tesista*** *(30-60 righe)*

1. **Revisione della letteratura**: Il tesista inizierà con lo studio della letteratura esistente sui Knowledge Graph per la cybersecurity in ambito Situation Awareness e sulle tecniche di estrazione di informazioni dai report di sicurezza.

2. **Implementazione di modelli di estrazione delle informazioni**: Il tesista implementerà un modello di Open Information Extraction (OIE) e utilizzerà un modello di Named Entity Recognition (NER) per estrarre entità e relazioni dai dataset di report APT.

3. **Costruzione e popolamento del Knowledge Graph**: lo studente integrerà le entità estratte nel Knowledge Graph, creando una struttura coerente e navigabile. Saranno utilizzate ontologie standard del dominio, come OWL e specifiche come, ad esempio, STIX.

4. **Sviluppo di un’interfaccia di interrogazione**: Lo studente svilupperà un’interfaccia che consenta agli utenti di esplorare il Knowledge Graph e recuperare informazioni pertinenti attraverso query SPARQL e Cypher. Sarà inoltre possibile formulare domande in linguaggio naturale, che verranno tradotte in SPARQL o Cypher, utilizzando modelli di large language per facilitare l’interrogazione. L'obiettivo è supportare la comprensione delle situazioni di sicurezza, aiutando l’utente a individuare o isolare informazioni chiave su potenziali minacce.

5. **Test e valutazione**: Il sistema sarà testato su dataset reali di report APT per valutarne l’efficacia e la precisione delle risposte. Saranno misurati parametri quali il tempo di risposta, la precisione delle informazioni recuperate e la robustezza del sistema nell'adattarsi a nuove minacce.

***Contributo ingegneristico/progettuale della proposta*** *(10-20 righe)*

Il contributo principale di questa tesi consiste nello sviluppo di un sistema per la gestione della conoscenza sottoforma di un Knowledge Graph per la Cyber SA al fine di identificare potenziali minacce attraverso l’estrazione dell’informazione raccolta nella KG.

Questo include:

* Costruzione del Knowledge Graph: Design di un framework che trasforma dati non strutturati in un grafo di conoscenza strutturato e interrogabile.
* Interrogazione del KG: Implementazione di strumenti per interrogare il Knowledge Graph in modo efficiente sia attraverso query SPAQL/Cypher che attraverso l’uso di domande in linguaggio naturale (e relativa traduzione in query degli opportuni linguaggi di interrogazione).
* Interfaccia utente: eventuale sviluppo di un’interfaccia che permette un accesso rapido e intuitivo alle informazioni di minaccia, supportando l'analisi di cyber threat in contesti aziendali o governativi.

**Data di approvazione** *(inserito dalla commissione)*: XXXXXX

**Periodo di validità** *(max 12 mesi)*:

*Inserire la data in cui la tesi non sarà più valida. Allo scadere di tale data, la tesi sarà cancellata e non sarà più disponibile.*

**Numero tesi disponibili** (max 2): 2

**Sede ERASMUS** *(laddove l’argomento sia nell’ambito di una convenzione ERASMUS, indicare la sede ospitante):*

**Note** *(opzionale):*